DOCUMENT RESUME

ED 218 328 TM 820 368

AUTHOR Trochim, William M.K.
TITLE Research Implementation.
PUB DATE [81]
NOTE 27p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Program Implementation; Research Design; *Research Methodology; Research Problems
IDENTIFIERS *Evaluation Problems; Implementation Analysis; *Research Implementation

ABSTRACT

Investigated is the topic of research implementation and how it can affect evaluation results. Even when evaluations are well planned, the obtained results can be misleading if the conscientiously-constructed research plan is not correctly implemented in practice. In virtually every research arena, one finds major difficulties in implementing the strategies which methodologists set forth. Some sources of implementation problems include the lack of an explicit research plan, inadequate training of staff, poor motivation of research participants, and intergroup conflicts. When devising an implementation plan, it is often useful to consider subparts of the research task separately and to determine for each part what problems are likely to occur. The following major research components should be considered: implementation of the program, sampling, the measures administered, the research design, data preparation, and data analysis. Quality control techniques, accounting techniques, pilot studies, planned replications, multi-trait multi-method approaches, monitoring and observation, design variations, statistical analyses, simulations, secondary analysis, or meta-evaluation can be used to assess and improve research implementation. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
W.M.K. Trochim
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED218328

Research Implementation

William M.K. Trochim
Cornell University

Consider the following hypothetical scene. Some agency - a school district, mental health center, welfare department - is required by law or regulation to evaluate one of its programs. Assume that the administrators, project personnel and evaluators are relatively knowledgeable and wish to do a conscientious job. They would be careful to specify the major program goals, to enunciate one or more central hypotheses and to define the key constructs operationally. They might even include multiple constituencies in the goal formulation process, especially clients, community groups and local politicians. They would be equally meticulous in their selection of a research methodology. Perhaps they would seek the advice of outside consultants or conduct a small evaluability assessment. When selecting a design they would try to account for the prominent threats to validity. They would be sure to consider the costs of various sampling, design and measurement alternatives. They would probably choose some combination of or standardized and tailored measures and would carefully specify how participants are to be selected for the program. If they are aware of relatively recent trends they would probably recognize that their program is likely to be degraded ot "watered down" in practice. At the very least they would expect that some program personnel will adhere to the specific procedures of the program more accurately and conscientiously than others. They may even incorporate this intuition into the evaluation itself by devising measures of program implementation and conducting site visits, audits and interviews of clients and program

TM 820 366

personnel. When conducting the statistical analysis they would consult textbooks and perhaps a statistician about the appropriate analysis (probably OLS). If the results indicate that the program participants gained on some measure (relative to non-participants) the administrators would most likely proclaim the program a success. If participants do not gain or, worse yet, lose ground on some measure, the crestfallen administrators would turn their thoughts to the devising of programs which do work.

Evaluations are not often planned as well as in this hypothetical example. Even when they are, the obtained results can be misleading if the conscientiously-constructed research plan is not correctly implemented in practice. This was the case in the New Jersey Income Maintenance experiment where high attrition rates threatened the randomness of group assignment and, ultimately, the interpretability of the results (Boeckman, 1978). In Head Start evaluation there has been a long-standing debate about the appropriate implementation of the statistical analysis (Campbell and Erlebacher, 1970; Cicirelli, 1970; Evans and Schiller, 1970). For at least a decade the Office of Economic Opportunity and its successor agency, the Community Services Administration, have been trying to devise a data collection and management information system which local agencies would be able and willing to implement (Lazar et al, 1981). Despite a well-designed evaluation plan, research in Title I compensatory education evaluation is plagued by an inability to adhere to such system requirements as the mandated assignment strategies and measurement schedules (Trochim, 1980). In virtually every research arena, including both laboratory and applied settings, one finds major difficulties in implementing the strategies which methodologists set forth.

Information about research implementation is difficult to locate in the literature. This is in part because so little formal research on implementation problems has been conducted. When such issues have been written about at all the discussions tend to be buried within larger research reports. To be sure, there are the general methodological exhortations to perform research well which are included in almost any introductory research text. There are also a number of recent articles discussing implementation of the treatment or program. In general, however, the literature on implementation questions is speculative in nature, after the fact, and tends to focus on only selected issues which are abstracted from the total research context.

This paper sets forth a slightly different view of research implementation issues. The emphasis is on prospective rather than retrospective consideration of implementation. Social research projects should routinely include an implementation plan which anticipates the major problems which are likely to occur and eliminates or minimizes them as much as possible. It is also important to consider the research project as a whole rather than focussing on specific problems only. Measurement difficulties can lead to design problems which in turn can cause analytic problems, and so on. The analysis of implementation should precede the initiation of a study and continue throughout the study's course. A wide variety of techniques are called for ranging from qualitative observation, case studies and interviews to quantitative side studies, simulations and secondary analyses. Finally, there needs to be a way to record and advertise major implementation problems so that previous mistakes are not repeated. From such information it should eventually be possible to identify the major sources of implementation difficulties and suggest

procedures for dealing with them.

The remainder of this paper is devoted to a discussion of these issues. First, some of the most likely causes of implementation problems are presented. Second, a tentative taxonomy of implementation problems which might provide a beginning point for the development of an implementation plan is discussed. Third, some strategies for detecting, minimizing, and if possible correcting such problems are presented. Finally, suggestions are given for how one might approach the study of the implementation of a particular research project.

Some Sources of Implementation Problems

Obviously, there are any number of reasons for why a research project might break down in practice. There is, however, almost no data on the causes of such breakdowns. In lieu of such information one is forced to speculate on likely sources of problems. The discussion which follows is suggestive of some of the potential sources of implementation difficulties which deserve empirical investigation and confirmation in the future.

Lack of an Explicit Research Plan

The key to good implementation is good planning. The planning scenario outlined in the introduction is seldom, if ever, attained in practice. When a specific plan is lacking, confusion results. Planning difficulties are often very subtle and difficult to anticipate. Consider what happened in one school district which was conducting an evaluation of its first grade compensatory education program (Visco, 1981). As is often the case, the evaluation proposal was written for the most part by the program administrators. They followed acceptable guidelines and selected the

regression-discontinuity research design. This design requires pretest and posttest data on both program and comparison students. Students must be selected for the program on the basis of a cutoff value on the pretest. In this case, the administrator who wrote the proposal selected the 24th percentile on an achievement pretest as the cutoff. Thus, all students scoring below the 24th percentile would be included in the compensatory program while those who score equal to or greater than that percentile would comprise the comparison group. Even with a selection criterion as explicit as this one implementation problems arose. The first question which occurred was whether the 24th percentile referred to the national or local test score distributions. Because this school district tended on average to be below the national norm, use of the 24th national percentile would result in far more program participants than use of the local percentile would. Another problem arose when the reading supervisor for the district argued that the 24th percentile criteria would only be appropriate for students who had gone to kindergarten. A different, more lenient cutoff should be used for non-kindergarten students. Matters got worse when the basic skills coordinator came along and said that he had told teachers to simply reserve ten program slots per school. His interpretation of the cutoff was that the ten lowest scoring students in each school should receive the program. The program supervisor countered this by arguing that such a procedure was illegal because it meant that different criteria would be applied to different schools. Ultimately the various parties did reach an agreement but, as is often the case, it was not a good one from an evaluation standpoint - students would be selected on a case-by-case basis.

This story illustrates how an apparently strong research design choice and explicit selection criterion were not sufficient largely because their implementation was not thought through carefully enough. The planning process involves at least two keys elements - the anticipation of future events and explicit agreement on how they will be managed. The obvious lesson is that these administrators should have reached agreement on the selection criterion before the research proposal was written.

Inadequate Training

Most research projects involve many people - administrators, program personnel, clients, evaluators, community members, and so on. Each of these groups performs different functions and will have to be trained in these tasks. Consider again the school district which would like to employ the regression-discontinuity design to evaluate its compensatory education programs. Even if the administrators reach agreement that the cutoff criterion should be the 24th local percentile, there are innumerable ways in which this could fail to be carried out. For example, there is the simple problem of communication. Local school administrators and classroom teachers would have to know what the cutoff score is and would have to have student scores in time to make a decision. While this might appear to be a relatively simple task it can often become logistically complex. If the students are entering the first grade, for example, tests will have to be corrected, and the scores forwarded to classroom teachers so that assignments can be made. This can lead to considerable delay in beginning the program. Furthermore, teachers who have not been informed of the importance of adhering to the assignment strategy or who are not convinced

of it may simply go ahead and choose the students they feel are most needy and begin the program rather than wait for the scores. In this case, it would be essential to good implementation that teachers and administrators be told why the chosen research design is being used, how the research is to be conducted, and convinced of the importance of adhering to the research strategy.

Poor Motivation of Research Participants

In many situations, evaluation of programs is seen as something which is forced upon staff members and clients. Often it involves the collection of considerable amounts of data and disrupts normal organizational procedures. Under these conditions it should not be surprising that there is resentment towards the evaluation and an unwillingness to cooperate with it. In order to insure that the research will be carried out as intended it is important to address motivational issues. Not much is known about how to improve the motivation of persons involved in a research project. Certainly, showing participants that there is a sensible research plan and that the information which is collected will be used is important. One often hears front-line staff complain that all their data collection efforts are wasted because no one ever looks at the information. In addition, potential implementation difficulties should be discussed with staff and sometimes with participants when formulating the research plan. This results in a plan which is stronger and to which the staff will feel a greater committment. Often, the question which the evaluation must address will not be the major question of interest to front-line staff or administrators. Despite the increased costs and effort, motivation may in this case be improved if the evaluation incorporates side studies to

address some of these other questions (Morell, 1979). Finally, it may often be helpful to provide incentives for participation in the research. These can range from financial payments to on-the-job benefits for staff members such as compensatory time off for time spent in gathering data.

Intergroup Conflicts

The example provided earlier illustrated how conflicts can arise between persons with different interests and responsibilities. In turn, these conflicts can have many sources. For example, research projects often compete for scarce funds with program areas. Sometimes having a research project conducted in a certain area will be perceived as a benefit, as when research funds make available more supplies, staff salaries, etc. In other cases the project will drain program funds by requiring extra staff time, revision of incompatible program procedures, and so on. Another source of conflict is related to ethical considerations. A case worker may perceive a person who is denied a program as truly needy and may resist the assignment procedure on moral grounds. Conflicts may also arise because the authority for managing the research is not clearly delineated. Local program administrators may override the research protocol, different agency offices can issue conflicting instructions and so on. The authority for making such decisions should be agreed upon by all involved parties when devising the research plan.

The tendency is to see these sources of implementation problems as obvious or mundane. Of course, there should be a good plan and conflicts should be resolved before hand. Despite general agreement on the importance of these problems we know almost nothing about how frequently

and in what settings they tend to occur. Furthermore, one seldom finds a research project where a conscientious attempt is made to anticipate and deal with such difficulties despite a more than ample history of implementation failures. The list above is merely preliminary. What is needed is a concerted effort on the part of methodologists and evaluators to systematically investigate these issues in order to develop a more useful formulation of threats to quality research.

A Taxonomy of Implementation Problems

Implementation problems can affect many aspects of a research project. When devising an implementation plan it is often useful to consider subparts of the research task separately and to determine for each part what problems are likely to occur. The following taxonomy suggests the major research components which might be considered in any given implementation plan.

Implementation of the Program

The central focus in any research is the program or treatment being studied. Most research designs assume that the program is a homogeneous entity, that it is being administered to each participant in exactly the same way. In recent years the reasonableness of this assumption has been seriously questioned. Two general types of problems tend to arise. First, the program may be delivered differently by different program administrators. This might be attributable to the idiosyncracies of the individuals or to selective misinterpretation of the program regulations. Second, even when there is relative homogeneity there can be systematic error in the manner in which the program is carried out. For example,

consider a study of a compensatory education program where students are supposed to be pulled out of their regular classroom in order to receive the special instruction. If one of the school districts which is participating in this hypothetical study finds the pull-out requirement distasteful, discriminating or incompatible with local procedures they might require that students be given the special instruction within their homeroom classes. Thus, while there would be relative homogeneity within the district, they would not be implementing the same program which was intended.

Implementation of the Sampling

Many research studies involve either systematic or random sampling of persons from a larger pool of potential research participants. Two major problems tend to occur here (Cochran, 1977). First, it is often extremely difficult to obtain a complete enumeration of the population in question. Second, even when a relatively complete enumeration is available it is usually difficult to obtain an acceptable response rate. Both of these problems introduce the possibility that the sample being studied is not representative of the target population of interest. Sampling theorist have attempted to examine these problems and devise strategies for dealing with them. Call backs are often used to increase response rates although there is reason to question the similarity between persons reached on the first try and those reached in subsequent tries. Even when persons are reached they may be reluctant to participate, especially if the information being requested is of a personal or sensitive nature. Guarantees of confidentiality of responses and the use of such techniques as randomized response methods may help to minimize such reluctance.

Implementation of the Measures

The process of collecting information for research purposes is complex and allows for many opportunities for error. One relevant factor is the setting in which the measurement takes place - some settings may be more conducive to good performance than others. This becomes especially troublesome in projects which have multiple measurement settings which can affect participants differently. When persons are to be tested in groups there is often the problem of absenteeism. While absenteeism on a single measure is troublesome enough the problem is exacerbated when multiple waves of measurement are needed. If, for example, a project involves the administration of a pretest and posttest, one can usually expect absenteeism on each measure (which cuts down on the sample size) and consequently, the number of cases for which matched data will be unavailable. Even respondents who are present for both measurements will often tend to leave certain question items blank. This can make it difficult to construct total scores for these persons and, as a result, their data will have to be omitted from the analysis. Inevitably, there are also errors in how respondents code answer sheets, especially when these are of the machine-readable type. Misunderstanding of the directions, miscoding of name, identification number and demographic information, and other problems of this nature, can invalidate an individual's data, further reducing sample size and possibly introducing bias into the estimate of program effect. In some settings, especially when standardized tests are used, make-up testing of respondents who were absent on the official test day can become problematic, especially if considerable time elapses between the original and make-up test dates.

Clearly, the loss of information due to poor implementation of the measurement process is a critical factor in much social research. Efforts directed to minimizing attrition and absenteeism, reducing deleterious setting factors, and increasing the understanding of the instructions are warranted.

Implementation of the Design

The topic of research design is often confusing to people who have never dealt with it before. The rationale behind the use of control groups, the need to follow a strict schedule, and other such requirements are often not immediately apparent. The first step in insuring that the research design is corectly carried out involves the communication of the rationale and mechanics of the design. The example presented earlier shows how confusion regarding design requirements can prevent the design from ever being implemented. Even if all partics involved have a clear idea of what is required there are often difficulties in carrying this out. This is especially true when dealing with strategies for assigning persons to various program conditions. Staff members will often be compelled through circumstance or conscience to allow exceptional cases into desireable conditions or out of undesireaable ones regardless of the explicit assignment rule.

Implementation of the Data Preparation

The term data preparation refers to anything which happens to the data between the time of measurement and the data analysis. This includes the collection and colation of data forms, the coding of data, computerized or manual scoring, computerized data entry and database construction, data

transformations, the selection of cases for analysis, and so on. This is probably one of the most underinvestigated areas of implementation despite the fact that serious problems can and do occur here. For example, data forms get lost, incorrect answer keys sometimes get used, data may be read int a computer using incorrect formats, and incorrect data transformations may be applied. Two of the more serious forms of difficulty involve matching and selecting cases for analysis. When multiple waves of measurement are involved, it is necessary to match the measures for individuals. When doing this on a computer, one must rely on some data item (the person's name or ID number) to make the match. Such a procedure can only be perfectly successful if the matching data item has been entered identically on each answer sheet. Thus, if a person changes his/her name between testings, uses the short form one time and the long form another (e.g., Bob versus Robert), uses a middle initial on one and not the other, or puts first and last names in incorrect fields, the match will not be successful. Using ID numbers is no easy solution to this problem. People forget their numbers, organizations accidentally assign the same number to two people, and so on. When selecting data for analysis, cases are often excluded for reasons which were not originally planned for. In a study of a compensatory education program, for example, students may be excluded from the analysis because of attrition, lack of a pre-post match, because they repeated the grade and/or repeated the program, because they moved to another classroom or school within the district, and so on. Again, these problems can both reduce sample size and introduce bias.

Implementation of the Analysis

The major implementation question concerns whether the correct analysis

is used correctly. This, then, has two parts. First, there is the selection of the correct analysis. In one sense this is a planning question. To the extent that a correct analysis exists for the data at hand, the decision to use it should be made before the study begins. On the other hand, someone ultimately has to select the analysis again when it is to be conducted. This may involve the apparently simple task of selecting a computer program which will conduct the analysis. Problems can arise here if the program must be selected from an assortment of packaged programs of if the person in charge of the analysis is unfamiliar with the relevant statistical jargon. Assuming that an acceptable statistical program is chosen, the next question is whether the program is used correctly. This requires considerable familiarity with the program in order to insure that procedures that handle missing data, index constructions, data formats, and the like, are being correctly implemented.

Methods for Assessing and Improving Implementation

The study of implementation encompasses a wide variety of topics and can be considered a legitimate, definable substantative area of its own. Because of this, virtually every available research method can be usefully applied to the investigation of implementation. The discussion which follows enumerates a number of research strategies or approaches which could be valuable for insuring that research is implemented correctly or failing that, for assessing the effects of poor implementation. The list is merely illustrative. Different categorizations of methods might be developed which are more appropriate to the task at hand. Also, no priority is implied by the order in which the topics are presented.

Quality Control Techniques

A number of strategies can be borrowed from industrial research methodology to help insure implementation. For example, it is common in industry when assessing the quality of materials produced to conduct periodic random spot checks. This idea can be extended to the social research context by randomly sampling records of participants during the course of a study to determine the accuracy of the information which is obtained. Similarly, one could sample days on which to observe the program being administered, sites at which measures are being taken, and so on. Prior to data analysis one could sample records from the final computerized database in order to track each data item from the time it was collected until it reached the final database, again to insure that the information is accurate. When implementation problems become apparent during a study one can sample problem cases in order to get a better fix on the source of the difficulty. Thus, spot checks of participants who are frequently absent or who often leave measurement items blank would be likely to provide some insight into the reasons for these problems.

An excellent example of the use of quality control techniques can be found in Crane and Maye (1980). They used Statistical Quality Control (SQC) techniques developed in industry to assess the frequency of "correctable" data errors in Title I compensatory education evaluations. Using a sample of the results reported to the State of Illinois for several years, they were able to demonstrate that previously undetected errors resulted in a positive bias in the State-aggregated estimate of program effect. Furthermore, their analysis indicated which of several types of errors were most prevalent and deserved the greatest attention in any data correction scheme. Finally, on the basis of their sample they were able to estimate the average time required to correct each error type and use these

time estimates in planning a cost-efficient data checking strategy.

Accounting Techniques

The accounting profession is devoted to assuring the accuracy of information and utilizes strategies which could be helpful in the study of implementation. Tabor (1977), for instance, provides an excellent discussion of evaluation data corruption from an accountant's perspective. He discusses a wide variety of corruption types, preventive methods, and ways of using audit procedures to detect data innacuracy.

A common accounting procedure which might be useful for assuring or assessing data quality is a double bookeeping strategy. Often, redundant information is collected throughout the course of a study. For example, almost all data which is collected includes the name or ID number of the person involved, perhaps along with other demographic information. Once this data is entered into a computerized database it is usually an easy matter to compare this information from multiple measures for each person to determine the accuracy of the data or whether the information has changed (e.g. the person has changed residence, marital status, religion)

Pilot Studies

All major research projects which are conducted should be preceded by a pilot study. Such a study should incorporate a significant qualitative or observational look at the process of conducting the research including interviews of research staff, program personnel and research participants. Alternative strategies for dealing with likely implementation problems can be tested and unanticipated problems discovered.

Planned Replications

Given sufficient resources, it is almost always advantageous to replicate the research project at several sites. This would help in the study of implementation if, for example, there was prior reason to expect that a specific implementation problem will occur with different frequency at the different sites. Thus, the problem itself becomes a factor in the study and its effect on the results can be assessed. For example, consider an agency where there are different sub-agencies. If the motivation of staff members can be assessed for each sub-agency, it would be possible to see whether initial motivation is related to the size of the program effect.

Multi-Trait Multi-Method Approaches

Within research sites it may be possible to assess implementation problems by including measures or research designs on which there is likely to be differential implementation. For example, rather than relying on a single measure of the construct of interest which may be too difficult, lengthy, hard to administer, and so on, it may be useful to include multiple measures of the same construct which differ on these characteristics. Different program effect estimates for each measure might then be attributable to differences in the implementation of the measures. The strategy could also be applied to research design. In a study of compensatory education, for example, side by side use of, say a randomized experiment and a regression-discontinuity design would yield two estimates of effect and could give some indication of whether implementation problems were occurring.

Monitoring and Observation

One of the most straightforward ways to assess implementation is to observe the project throughout its course. While this may appear obvious, such activities are routinely neglected in social research. An implementation plan might include, for instance, the systematic observation of data collection, routine monitoring of the program, spot checks of intake procedures, and so on.

Design Variations

One example of a design strategy which provides information about implementation is the use of double pretests (Boruch, 1977). Here, instead of relying on only a single preprogram measure and posttest, two pretests are given (ideally spaced the same time apart as the second pretest and posttest). Many implementation problems which would occur during the course of a study can be detected from pretest 1 to pretest 2 and assessed through a dry run analysis. Thus, significant program effects which occur between the two pretests may indicate implementation problems which lead to biased results.

Another useful approach involves "coupling" two or more designs within the same research project, thereby creating a new, hybrid design (Boruch, 1975; Trochim, 1980). For example, suppose that in a particular multi-site study two research designs, a randomized experiment and a non-equivalent group design, were both implemented at different sites. While the randomized experiment should under ideal conditions lead to less equivocal statements about the effect of the program, the randomization procedure often makes it more difficult to implement than its non-equivalent

counterpart. Use of both designs simultaneously should provide some indication of whether implementation problems may be occurring.

Statistical Analyses

Often it is possible to explore the data statistically to see whether implementation problems may be affecting estimates of program effect. For example, a plot of the frequency of a problem (e.g. attrition, missing data) against pretest values would tell whether the problem is distributed evenly across the pretest. If an implementation problem is suspected within a particular range of data it may be useful to conduct multiple statistical analyses including and excluding the relevant data to see whether estimates of program effect are different. Thus, depending on the research design, if floor effects are thought to exist on some measure, it may be valuable to analyze the complete dataset and a subset which excludes data within the likely floor range to see if the results differ.

One example of an interesting analytic strategy for detecting implementation problems was suggested in Jurs and Glass (1971). To investigate the effects of attrition they construct a 2 x 2 ANOVA (or MANOVA) where the dependent measure is some pre-program measure (e.g., a pretest, demographic variable). The first factor is program (i.e., program versus control group) and the second is attrition (i.e., attrited versus complete group). Tests for main and interaction effects yield information about whether the attrition is random relative to the dependent variable in the analysis. In addition, such an analysis also enables a determination of whether the attrition has threatened internal validity, external validity, or both. Boeckman (1978) used this procedure to conclude that attrition rates affected the results of the New Jersey Negative Income Tax

Experiment. The procedure is easily extended to other implementation problems such as data coding errors and data exclusions.

Simulations

The advantage of simulation is that it allows the researcher to manipulate the parameters of a problem and then see how the results would be affected. For example, if attrition is likely to be more dominant in the program group than the comparison group, one can often construct a computer simulation using various attrition distributions and determine how the results will change. In general, simulations of this sort are useful in the planning stage when trying to decide whether to proceed with a study or when investigating alternative analytic approaches to likely implementation problems. Real-life simulation might be used in the pilot study or incorporated into the actual study by using confederates of the researcher to pose as clients. The researcher can control the information which the confederates supply and use their data to check on data manipulation accuracy and their insights to determine likely implementation problems from the participant's point of view.

Secondary Analysis

Different researchers tend to approach the same research task with different analytic perspectives. The use of multiple independent analyses of the same data can act as a check on the implementation of the original analysis (Boruch and Wortman, 1978). This technique is similar to the notion of double bookkeeping - discrepancies in results may be suggestive of implementation difficulties.

As an example of how secondary analysis may be useful in the study of implementation, Trochim (1980) reanalyzed data from several school district's compensatory education programs and showed that the results were, to some extent, dependent on different strategies for matching pre and posttest scores, attrition rates, misassignment of participants, and so on.

Meta-Evaluation

There will be times when implementation problems are difficult to detect within a single research study. When the program in question is instituted in the same manner in several settings discrepancies in results may be indicative of implementation problems. Sometimes two or more similar settings may utilize different measures, designs, analyses, and so on. Here, comparisons of program effect across settings may incorporate these variations directly within a meta-analysis (Cook and Gruder, 1979) framework. For example, Trochim (in press) compared two commonly used research designs across several hundred compensatory education evaluations. The analysis provided evidence of a consistent bias in one of the designs and pervasive implementation difficulties in both.

Discussion

This paper has outlined some of the relevant dimensions of research implementation and its assessment. It is appropriate at this point to suggest how the implementation task might have been approached by the hypothetical agency outlined in the introduction.

A first step in the implementation planning process would be for the major persons involved in the study to enumerate the most likely problems

which can be expected. The agency administrators, program administrators and staff, and evaluation personnel presumably have a wealth of personal experience from which to draw in making this initial determination. Evaluation staff might then attempt to devise strategies which minimize the problems which are expected. A record keeping system could be constructed to anticipate attrition, mobility, and so on. Training procedures could be set up to help convey confusing aspects of the project to staff and participants. Multiple designs, measures and sites which are likely to exhibit different degrees of implementation accuracy can be incorporated into the plan. Procedures for providing continuous feedback to line staff on quality of data and adherance to procedure could be developed. With this enhanced research plan it would be possible to conduct a pilot study to see how well the procedures work and to detect previously unanticipated problems. The pilot study can be observed carefully and followed up with interviews of involved parties including staff and research participants. At this point it should be possible to devise a final research plan which is much less naive to the problems of executing the research. Once this project is in progress, ongoing monitoring and feedback would be essential. This should include periodic spot checks of the measurement process and of the program. On the basis of the pilot study there probably will be some design variation and multiple measurements. When problems are detected they will be dealt with as quickly as possible. For example, a particularly high rate of absenteeism at one research site should be examined immediately to determine whether it results from reaction to the program or research, sickness, harvest season, or other factors. When the data is processed it will be necessary to make sure that it is handled correctly. The agency might have two independent groups simultaneously

process the data or might conduct spot checks. If tests are to be computer scored, a certain percentage could also be scored by hand. If this is done for the pretest, problems which might occur on the posttest can be anticipated and planned for. When all of the data has been assembled and stored in computer files it can be given to several independent analysts who can be instructed to investigate specific implementation problems (e.g. by plotting the frequency of the problem against pretest scores) and to conduct the final analysis using the statistics which they deem appropriate. The final results can then be presented to persons involved with the research and their impressions of the sensibility and accuracy of the results can be assessed.

The scenario outlined above is merely suggestive. Obviously, more accurate information about implementation problems is desireable. We know very little about the frequency with which various problems occur in different research contexts. We know even less about the effects which the problems have on the results. Furthermore, it will often be unreasonable to expect that in any given research project an attempt should be made to deal with all possible problems. Instead, there is a need for information and procedures to make possible a prioritization of the problems so that those which are likely to be most serious can be given the most attention. It is also not clear how much of the total research effort should be devoted to insuring implementation. Given the inevitable limitations on resources available for research, a decision to devote more time and effort to implementation will almost surely limit what is available for other aspects of a project.

What is needed, then, is more systematic study of research

implementation. Experience has shown that any number of problems can ruin even the best made research plan. Until procedures are developed which can assure a reasonable degree of implementation the knowledge claims of social science researchers will often continue to be justifiably suspect.

References

Boeckman, M.E., Attrition: A threat to validity in randomized field experiments. Unpublished draft manuscript, The George Washington University, May, 1978.

Boruch, R. F., Coupling randomized experiments and approximations to experiments in social program evaluation. Sociological Methods and Research, 4, 1, 31 - 53, August, 1975.

Boruch, R. F., Double pretests for checking certain threats to the validity of some conventional evaluation designs, or Stalking the null hypothesis. Unpublished manuscript, Northwestern University, Evanston, Illinois, 1977.

Boruch, R. F. and Gomez, H., Power theory in social program evaluations. In L. Tornatsky (Ed.) Program Evaluation. New York: Pergamon, 1979.

Boruch, R. F. and Wortman, P. M., An illustrative project on secondary analysis. New Directions for Program Evaluation. 4, 1978, 89 - 110.

Campbell, D. T. and Boruch, R. F., Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In C. A. Bennett and A. A. Lumsdaine (Eds.) Evaluation and Experiment. New York: Academic Press, 1975.

Campbell, D. T. and Erlebacher, A., How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.) Compensatory Education: A National Debate. New York: Brunner/Mazel, 1970.

Cicirelli, V. G., The relevance of the regression artifact problem to the Ohio-Westinghouse evaluation of Head Start: A reply to Campbell and Erlebacher. In J. Hellmuth (Ed.) Compensatory Education: A National Debate. New York: Brunner/Mazel, 1970.

Cochran, W. G., Sampling Techniques. 3rd edition, John Wiley and Sons: New York, 1977.

Connor, R. F., Selecting a control group: An analysis of the randomization process in twelve social reform programs. Evaluation Quarterly, 1, 2, 195 - 244, 1977.

Cook, T. D. and Gruder, C. Metaevaluation research. In L. Seechrest, S. West, M. A. Phillips, R. Redner and W. Yeaton (Eds.) Evaluation Studies Review Annual, Volume 4. Sage Publications: Beverly Hills, California, 1979.

Crane, L. R. and Maye, R. O., Effects of correctable errors on Title I NCE gain estimates and implications for statistical quality control. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, Massachusetts, April 7 - 11, 1980.

Evans, J. W. and Schiller, J., How preoccupation with possible regression artifacts can lead to a faulty strategy for the evaluation of social action programs: A reply to Campbell and Erlebacher. In J. Hellmuth (Ed.) Compensatory Education: A National Debate. New York: Brunner/Mazel, 1970.

Glass, G. V., Primary, secondary and meta-analysis of research. The Educational Researcher, 10, 3 - 8, 1976.

Jurs, S. G. and Glass, G. V., The effect of experimental mortality on the internal and external validity of the randomized comparative experiment. The Journal of Experimental Education, 40, 1, 62 - 66, 1971.

Lazar, K. S. Zeisel, L. and Klein, R., Studies in Community Action, Volume I: A Legislative and Administrative History. The Foundation for Human Service Studies, Ithaca, New York, September, 1981.

Leithwood, K. A. and Montgomery, D. J., Evaluating program implementation. Evaluation Review, 4, 2, 193 - 214, 1980.

Morell, J. A., Program Evaluation in Social Research. Pergamon Press: New York, 1979.

Tabor, J.G., The role of the accountant in preventing and detecting information abuses in social program evaluation. In H. W. Melton and D. J. H. Watson (Eds.) Interdisciplinary Dimensions of Accounting for Social Goals and Social Organizations. Grid, Inc.: Columbus, Ohio, 1977.

Trochim, W. M., The regression-discontinuity design in Title I Evaluation. Ph.D. Dissertation, Northwestern University, 1980.

Trochim, W. M., Different designs, similar programs, different results: Methodologically-based discrepancies in the results of compensatory education evaluations, Evaluation Review, in press.

Visco, R., Personal communication, May, 1981.

Williams, W., Implementation analysis and assessment. In T. D. Cook, M. L. DelRosario, K. M. Hennigan, M. M. Mark and W. Trochim (Eds.) Evaluation Studies Review Annual, Volume 3, Sage Publications: Beverly Hills, California, 1978.